SUMMARY REPORT OF DATA STANDARD FOR BIOLOGICAL TAXONOMY

CONTRACT NO. 68-WI-0055
DELIVERY ORDER NO. 057

Prepared for:

United States Environmental Protection Agency
Office of Information Resources Management
401 M Street, SW.
Washington, DC 20460

Delivery Order Project Officer:

Geoffrey C. Steele

Prepared by:

EPA Systems Development Center
(A Contractor Operated Facility)
Science Applications International Corporation
200 North Glebe, Suite 300
Arlington, VA 22203

CONTENTS

1.0   INTRODUCTION

The Environmental Protection Agency (EPA), Office of Information Resources Management (OIRM), Enterprise Information Management Division (EIMD), through the mission of its Information and Data Management (IDM) Program, has introduced the concept of centralized management and coordination of EPA's distributed information and data resources. The data policies and standards of the IDM program will be supported by data architecture, data models, standard data element domains, and a repository of standard data elements. These work products will facilitate data sharing efforts and effective management of EPA's information resources. The focus of this report is the data elements and domains for proposed standard data elements for biological taxonomy.

1.1   Background

The One Stop Reporting Program was initiated by the EPA Executive Steering Committee (ESC) for the purposes of improving the Agency's management of environmental information and reducing the burden of reporting from the regulated community. To achieve the purpose of improving information management the One Stop team has involved the stakeholders, particularly the states, in identifying the need for data standards and in participating in their development. To further the goals of the One Stop Program, the Information and Data Management Service Center (IDMSC), Delivery Order (DO) 57, has been tasked to identify needed data standards that will facilitate data sharing and to develop these data standards.

The Reinventing Environmental Information (REI) Program recently undertaken by the Agency has also identified the implementation of data standards and the development of electronic reporting processes as major goals. The EPA REI Action Plan identified the biological taxonomy data standard as one of the six priority standards to be published this year. The standard is scheduled to be adopted by EPA under the REI Program. The draft data standard will be reviewed by the EPA Biological Identification Work Group. The Environmental Data Registry (EDR) will be the vehicle for review and approval of the biological identification group of data elements, and will be the ultimate repository of the Agency approved data standards. These standard elements will be used by program systems that record identification information for biological entities.

Taxonomy is the discipline within the biological sciences concerned with the description and naming of the species of the world, as well as the genetic and evolutionary relationships among these species. It is the science that defines and documents biological diversity. In short, taxonomic data and information are necessary to support all types of biological inventory, monitoring, and research. Further, the ability to refer to standardized taxonomic nomenclature is a prerequisite for biological data sharing and comparison among different agencies and

organizations. Standardized names of organisms allow users to look at synonyms or alternative names that have been used to describe the same species in different geographic regions or at different times.

1.2   Purpose

The purpose of the data standard is to promote the collection of biological identification data in a standard fashion so that it can be employed consistently in monitoring programs. Standards also promote the reuse of information, which can save the considerable expense of developing crosswalk information necessary to exchange information about biological entities.

The purpose of this report is to propose standard data elements for biological identification information, and to specify domains for those data elements.

1.3   Scope

This document is limited to proposing data elements for inclusion in EPA information systems that contain information on biological entities.

1.4   References

The following publications were used as references for preparing this report:

- *Biological Data Profile of Content Standard for Digital Geospatial Metadata,* Federal Geographic Data Committee (FGDC).

- *Content Standard for Digital Geospatial Metadata,* FGDC.

- Executive Order 12906, *Coordinating Geographic Data Acquisition and Access: The National Spatial Data Infrastructure*, April 11, 1994, President William Clinton.

- *List of EPA Information Systems Containing Biological Taxonomic Data*, SDC-0055-057-KG-7030, March 31, 1998.

- *Data Field Definitions*, Multi-Jurisdictional Database, The Nature Conservancy, March 24, 1998.

System Documentation for various Agency information systems were included in this review.

The following World Wide Web (WWW) sites were used as references for this report:

- FGDC Home Page, http://www.fgdc.gov/.

- National Biological Information Infrastructure (NBII) Home Page, http://www.nbii.gov/.

- Integrated Taxonomic Information System (ITIS) Home Page, http://www.itis.usda.gov/itis/.

- The Universal Virus Database, Research School of Biological Sciences, The Australian National University, http://life.anu.edu.au/viruses/welcome.htm.

2.0   METHOD

Early in 1998 a review of Agency information systems was done to determine which systems contained biological taxonomy information and the type of information that was being stored. The working paper, *List of EPA Information Systems Containing Biological Taxonomic Data,* was the first step towards the development of the Biological Taxonomic Data Standard. The report provided not only information about those EPA systems that contain some biological data but also identified EPA staff and others who have knowledge about the use of biological taxonomy information in the Agency and who might serve as members of a work group on a Biological Identification Standard. Appendix A is an updated version of the table that accompanied the working paper.

The information systems identified in the report were reviewed and a crosswalk of data elements used for the identification of biological entities in the systems was prepared (Appendix B). Not every system that was noted in the Appendix A table actually recorded biological identification information. In addition, a review of other biological identification standardization efforts was also done, including the work of the FGDC Biological Data Working Group, the NBII (Appendix C), and the ITIS (Appendix D). The relevant data elements from this review are included in Appendix B.

The EPA Biological Identification Work Group met several times to review the types of biological information that was currently used in EPA information systems. The group reviewed the Nature Conservancy Multi-Jurisdictional Database project, the FGDC efforts, the NBII project, ITIS, and other program system projects, such as the STOrage and RETrieval (STORET) System and the Office of Pesticides information systems.

After the review was completed, a proposed standard set of data elements for the identification of biological entities was prepared (Appendix E).

3.0   RELATED DATA STANDARDIZATION EFFORTS

One of the major challenges involved in the collection, exchange, and use of biological information is the lack of ready access to standardized information on the names and taxonomy of organisms. There are many different systems and approaches within the biological science community on how identified species should be classified and named. This has led to the relatively common situation in which the same species may be referred to under different names. Understandably, agencies and institutions that have made significant investments in gathering species-specific information according to one naming convention are reluctant to revise their existing information to adhere to another naming system. Without standardized information, it is difficult to compare, combine, and exchange biological information among such different distributed information sources. At the same time, the verification and tracking of species names that is required to develop this type of standardized authority is a highly resource-intensive enterprise that individual agencies working on their own have not been able to support.

The following groups have been involved in the development of data standards for the use of biological identification information. Each has a slightly different perspective on the type of information that is being addressed, but the need to identify biological entities is common.

3.1   Federal Geographic Data Committee Standards

Executive Order 12906, signed in April 1994, called for establishment of the National Spatial Data Infrastructure (NSDI) to promote the sharing of geospatial data throughout all levels of government, the private and non-profit sectors, and the academic community. Section 3, paragraph (b) states: "Standardized Documentation of Data, … each agency shall document all new geospatial data it collects or produces, either directly or indirectly, using the standard under development by the FGDC, and make that standardized documentation electronically accessible to the Clearinghouse network." The NSDI encompasses policies, standards, and procedures for organizations to cooperatively produce and share geographic data. The 16 federal agencies that make up the FGDC are developing the NSDI in cooperation with organizations from state, local and tribal governments, the academic community, the private sector, and the international community.

In 1994, the FGDC approved the *Content Standard for Digital Geospatial Metadata* (Version 1) that set standards for the geospatial metadata. This was updated to Version 2 in 1998. The standard was developed from the perspective of defining the information required by a prospective user to determine the availability of a set of geospatial-based data, to evaluate the fitness of the set of geospatial-based data for an intended use, to determine the means of accessing the set of geospatial-based data, and to successfully transfer the set of geospatial-based data. Although the FGDC metadata content standard provides excellent documentation of a data set from the geospatial perspective, it is limited and, in some aspects, inadequate, for describing data from the biological science perspective.

An additional section to expand the scope of the existing standard to include biological information was supported by the FGDC. The *Biological Data Profile of the Content Standard for Digital Geospatial Metadata* was developed by the United States Geological Survey (USGS), Biological Resources Division (BRD). The USGS/BRD has played a major role in the development of a distribution framework for biological data and information sources called the NBII, which serves as the basis for the *Biological Data Profile*. A key element in the development of a distributed federation of biological data and information is the availability of a standardized approach to metadata documentation that is adapted for use by the biological science community. The proposed *Biological Data Profile of the Content Standard for Digital Geospatial Metadata* standard can be used to specify metadata content for the full range of biological resources data and information. This includes biological data, which are explicitly geospatial in nature (such as monitoring or sampling information), as well as data that are not explicitly geospatial (such as data resulting from laboratory-based research). The standard also supports other information sources, such as research reports, field notes, or specimen collections. This standard is intended to support increased access to and use of biological data among users on a national (and international) basis and to help to broaden the understanding and implementation of the FGDC metadata content standard within the biological resources community.

Appendix C lists the biological identification data elements that are part of the proposed *Biological Data Profile*.

3.2   National Biological Information Infrastructure

The NBII is an electronic gateway to biological data and information maintained by federal, state, and local government agencies; private sector organizations; and other partners around the nation and the world. The NBII, led by the USGS, is dedicated to the development of an electronic federation of biological data and information sources. Its success rests on a growing network of partners who want to share biological information. The goal of the NBII is to provide swift user access to biological databases, information products, directories, and guides maintained by federal, state, and local government agencies, non-government institutions, and

private sector organizations in the United States (U.S.) and around the world. To meet this goal, the NBII is developing a distributed collection of biological information sources, together with the necessary tools to help users find the biological information they need, combine information from different sources, and apply information to natural resources management decisions. The types of biological information addressed within the framework of the NBII includes:

- Information at the genetic, cellular, microbiological, anatomical, physiological, organism, and ecological levels.
- Reference information on biosystematics and nomenclature.
- Results of laboratory-based research on the causes of wildlife diseases.
- Information on specimens in natural history museums.
- Directory-type information on biological experts and specialists.
- Bibliographies, publications, and reports.

A key element in fostering development of the NBII as a distributed mechanism for biological data and information is the availability of a standardized format to use in describing the data and information. This structure can help people quickly and easily compare and contrast among many different sources to choose those that best meet their needs. A metadata standard for biological information is a fundamental component of the overall NBII effort.

The NBII biological metadata standard can be used by anyone interested in increasing access to biological data or information. This includes explicitly geospatial biological data, biological data that are not explicitly geospatial, and biological resources information. The biological metadata standard should be hardware and software independent. Any metadata created in compliance with this profile should function in harmony with metadata created with the FGDC geospatial metadata standard and with associated metadata search, indexing, and exchange tools, formats, and protocols.

Appendix C lists the biological identification data elements that are part of the proposed *Biological Data Profile*. This Appendix contains only the NBII data elements that are found in the proposed standard.

3.3   Integrated Taxonomic Information System

ITIS is a partnership of United States, Canadian, and Mexican agencies, other organizations, and taxonomic specialists cooperating on the development of an on-line, scientifically credible, list of biological names focusing on the biota of North America. ITIS is also a participating member of Species 2000, an international project indexing the world's known species. Six federal agencies are working together to foster and modernize the system for naming nature's living organisms:

- United States Geological Survey.
- Environmental Protection Agency.
- National Oceanic and Atmospheric Administration.
- Natural Resources Conservation Service.
- Agricultural Research Service.
- Smithsonian Institution's National Museum of Natural History.

Each partnering agency has a mission to inventory, monitor, research, or manage biological resources. This creates a common need for a vocabulary of shared taxonomy. Taxonomic nomenclature provides the most fundamental building block for information sharing on biological resources - the scientific name. The information included in ITIS conforms to existing nomenclature standards including the International Code of Botanical Nomenclature and the International Code of Zoological Nomenclature. A five-kingdom system has been adopted as a standard. The kingdoms are Monera, Protista, Plantae, Fungi, and Animalia.

Several existing biological information databases are being included in ITIS:

- The USGS/BRD Survey Project has developed a Checklist of Vertebrates of the United States, the U.S. Territories, and Canada. An updated version of the original Checklist, to which freshwater fishes have been added, will be included in ITIS.

- The National Oceanographic Data Center (NODC) Taxonomic Code is a system of numerical codes used to represent the scientific names of organisms chiefly in support of archiving oceanographic data.   As part of the original agreement among the participating agencies, updated and reviewed NODC data will be incorporated into ITIS.

- The PLANTS database, managed by the Natural Resources Conservation Service (NRCS) of the Department of Agriculture, will be included in ITIS. The database includes all native or naturalized vascular plants, mosses, lichen, liverworts, and hornworts known to occur in the United States.

The ITIS project is actively identifying potential contributors to determine the availability of data and expertise for ITIS. All ITIS data contributors are specialists in their particular taxonomic areas. The contributors listed below have either compiled data, coordinated the compilation of data, are currently preparing data for ITIS, or will be serving as data stewards for an information area:

- **Algae** - A hierarchy of all algae and a complete list of freshwater North American Diatoms has been prepared by Pat Kociolek, California Academy of Sciences, San Francisco, California.

- **Amphibians of North America** - Roy McDiarmid, USGS, Washington, DC is the ITIS steward for amphibians.

- **Birds of the World** - Richard Banks, USGS, Washington, DC and Alan Peterson, Walla Walla, Washington are the ITIS stewards for birds.

- **Cephalopoda of the World** - Mike Sweeney and Clyde Roper, Smithsonian Institution National Museum of Natural History, and Mike Vecchione, National Marine Fisheries Service, Washington, DC. Mike Sweeney is serving as the ITIS steward for Cephalopods.

- **Coleoptera of North America** - Margaret Thayer, Field Museum of Natural History, Chicago, Illinois, is the ITIS steward and is coordinating the work of many Coleopterists to assemble the names for this large group.

- **Decapoda of North America** - Austin Williams, National Marine Fisheries Service, Washington, DC is the ITIS steward for Decapods.

- **Diptera of the World** - Chris Thompson, Systematic Entomology Laboratory, United States Department of Agriculture (USDA), Washington, DC is the ITIS steward and is coordinating with Neal Evenhuis, Bishop Museum, Honolulu, HI, the work of many dipterists to assemble the names of this large group.

- **Fish of North America** - Wayne Starnes, North Carolina State Museum of Natural Science and Bruce Collette, National Marine Fisheries Service, Washington, DC are the ITIS stewards for fish.

- **Isopoda of the World** - Brian Kensley and Marilyn Schotte, Smithsonian Institution National Museum of Natural History. Brian is the ITIS steward for Isopods.

- **Mammals of North America** - Alfred Gardner, USGS, Washington, DC is the ITIS steward for mammals.

- **Mollusca of the World** - Paul Scott, Santa Barbara Museum of Natural History, is the ITIS steward for molluscs.

- **Odonata of the World** - Data come from Rosser Garrision, Azusa, CA.

- **Reptiles of North America** - Roy McDiarmid, USGS, Washington, DC is the ITIS steward for reptiles.

- **Trichoptera of the World** - John Morse, Clemson University, will prepare the Trichoptera World Checklist for ITIS. John is the ITIS steward for caddisflies.

- **Tunicata of the World** - Linda Cole, Smithsonian Institution National Museum of Natural History, is the ITIS steward. Work is complete and data are being converted for North American tunicates and is underway for species of the Caribbean.

- **Turbellaria of the World** - Seth Tyler, University of Maine, Orono, has updated his list for ITIS. It is in the process of conversion for the ITIS database. Masaharu Kawakatsu and Robert Ogren have also participated in these efforts.

- **Vascular and Non-vascular Plants of North America** - Scott Peterson, USDA Natural Resources Conservation Service, National Plant Data Center, Baton Rouge, LA, is the ITIS steward for the PLANTS database. The taxonomic data within PLANTS is derived from cooperative efforts by John Kartesz, BONAP, North Carolina Botanical Garden (vascular plants); Robert Egan, University of Nebraska-Omaha (lichens); Marshall Crosby, Missouri Botanical Garden (mosses); and Ray Stotler and Barbara Crandell-Stotler, Southern Illinois University-Carbondale (liverworts and hornworts).

As ITIS contributors complete review or compilation of data, those groups will be certified as having met ITIS standards and will be loaded into ITIS. The ITIS Home Page is updated periodically with status information on recently updated information. ITIS does not include viruses and bacteria.

A data model of ITIS is available on the ITIS Home Page. Appendix D is a list of biological identification data elements found in ITIS.

4.0   BIOLOGICAL TAXONOMIC DATA ELEMENTS USED IN EPA AND OTHER INFORMATION SYSTEMS

The information systems listed in Appendix A were reviewed for data elements used to identify biological entities. Appendix B illustrates the type of information that was found across several systems. Most systems record either a systematic and/or a vernacular name of the biological entity of interest. This is generally done by some combination of Kingdom, Division-Phylum, Class, Order, Family, Genus, and Species. Not all systems record each level of information; at minimum a binomial name is used (Genus and Species). Some systems have separate records for each part of the biological taxonomic name but others record the information in a single data element.

There are no current EPA standards for either the taxonomic name or the vernacular name, (although the Agency fully supports the work that is being done in the ITIS project and a related Multi-Jurisdictional database being prepared by the Nature Conservancy). The nomenclature used can vary by geographic location, researcher, or program system. Because of these variables, comparison of information across information systems is difficult, at best. The use of data standards for identification of biological entities would facilitate information sharing.

5.0 PROPOSED DATA ELEMENTS FOR BIOLOGICAL TAXONOMY INFORMATION

Appendix E is a list of proposed standard data elements. The FGDC/NBII draft biological data standard addresses metadata for sets or collections of data while the EPA biological identification standard addresses data elements designed to uniquely identify biological entities in EPA information systems. The majority of the identification type data elements are found in both ITIS and the FGDC/NBII standard. Some additional data elements, such as Program System Biological Name and Context, and Biological Group Descriptive Text, have been added as a result of discussions of the Biological Identification Work Group. These data elements will be added to the EDR to facilitate review.

The review of information systems that utilize biological information and of the ongoing biological identification projects of several organizations formed the basis for the conclusion that the ITIS project most closely matches the needs of the EPA for biological taxonomy information. ITIS provides a unique taxonomic serial number for each taxon, records extensive taxonomic information, includes vernacular or common names, maintains source information, and has good quality, reviewed data. Using the ITIS taxonomic serial number, Agency systems could link to ITIS records to retrieve information. ITIS is continuing to add new taxonomic coverages as noted in Section 3.3. Upon request, new information could be researched and added to meet specific project needs. For topics of interest to Agency programs, such as viruses or bacteria, that are not currently planned for inclusion in ITIS, a separate information source could be maintained and accessed via a cross-linked EPA Biological Identification Number (EPA BIN). For example, The Universal Virus Database at

the Research School of Biological Sciences, The Australian National University is an authoritative source of virus nomenclature. This database uses a decimal numbering system (which can be represented by a eight digit number) to express taxonomic relationships and this number could be linked to an EPA BIN.

The minimum information that would be stored in a local system would be either the ITIS Taxonomic Serial Number (TSN) or the EPA BIN. An EPA BIN would be assigned to biological entities that are not, currently, part of ITIS. With either of these, taxonomic information could be retrieved as needed from ITIS or other database sources. The ITIS TSN to be recorded should be either the number associated with the binomial (Genus and species) or trinomial name (Genus, species, and subspecies).

If it is necessary to store systematic name information in a local database, it could be recorded as a single field or could be broken into parts to store the classification levels in separate fields. (ITIS provides the classification levels as separate fields; other programs often have the name as one record.) A Systematic Name could consist of a binomial or trinomial name, such as *homo sapiens*, or if desired, a more complex name including Kingdom, Phylum, Family, Genus, Species, or more detailed nomenclature information could be stored. ITIS would be the source of the name and also could provide source information (context) for each name. Common or vernacular names, synonyms, and program systems names are also meaningful for identification and communication. The context of these names can be useful in determining where a particular name is utilized. ITIS includes vernacular name information. For regulated substances, citation information could include information about documentation that establishes the requirement for monitoring a biological entity. Group information can be used to relate taxa. Neither citation information nor group information is part of ITIS.

The following data elements illustrate the type of information that should be collected about a biological entity. Appendix E shows specific data elements and definitions that have been recorded in the EDR. All program systems that identify biological entities must use an ITIS TSN to link to nomenclature and identification. If an ITIS TSN is not available, the EPA BIN can be used. All other data elements are optional.

- **ITIS Taxonomic Serial Number (ITIS TSN)** – The unique serial number assigned by ITIS to a biological entity. If a biological entity does not have an ITIS TSN, a request can be made to have one assigned or an EPA Biological Identification Number can be used. Either a TSN for a binomial or trinomial name can be recorded. (Mandatory)

- **EPA Biological Identification Number (EPA BIN)** – A unique, non-intelligent number assigned to a biological entity. This will be a unique number assigned to each biological entity and could provide a link to information available in other databases. (Conditional)

- **Systematic Name** – The name assigned to a biological entity by a classification identification system. This could be recorded in one data element or broken down into its components such as:

- Kingdom –The name that describes the portion of the systematic name that represents the Kingdom.

- Division/Phylum – The name that describes the portion of the systematic name that represents the Division/Phylum.

- Class – The name that describes the portion of the systematic name that represents the Class.

- Order – The name that describes the portion of the systematic name that represents the Order.

- Family – The name that describes the portion of the systematic name that represents the Family.

- Genus – The name that describes the portion of the systematic name that represents the Genus.

- Species – The name that describes the portion of the systematic name that represents the Species.

Additional divisions such as super- and sub- categories of this list could be used where required. ITIS provides such information. Data elements to accommodate these additional categories have been included in Appendix E.

- **Systematic Name Context** – The classification system used to assign the systematic name to a biological entity.

- **Vernacular Name** – A common name associated with an occurrence of a biological entity.

- **Vernacular Name Context** – The source of a common name.

- **Synonym Name** – One of two or more names applied to a single biological entity.

- **Synonym Name Context** – The source of a synonym name.

- **Program System Biological Name** – The name applied to a biological entity in a program system.

- **Program System Biological Name Context** – The program system that is the source of a biological name.

- **Group** – Any collection of biological entities that can be related in some way.

- **Group Description** - Text describing the composition or origin of a grouping of related taxa.

- **Citations** – References to the Code of Federal Regulations or other documentation that establishes the requirement for monitoring a biological entity.

APPENDIX A

EPA Information Systems Containing Biological Data

EPA Information Systems Containing Biological Data

System	Application Type	Abstract	Contact Information
AQUIRE	Database	Aquatic Toxicity Information Retrieval System (AQUIRE) contains information on the toxic effects of chemicals to aquatic organisms and plants. Toxicity test results and related testing information for any individual chemical from laboratory and field aquatic toxicity tests are contained in AQUIRE. Acute, sub-lethal and bioconcentration effects are recorded for fresh water and marine organisms.	ORD. Christine Russom (218) 529-5000
BASINS	Database/Modeling	Better Assessment Science Integrating Point and Non-point Sources (BASINS) is a new tool which integrates environmental data, GIS (ArcView) and modeling tools on your PC. BASINS will integrate data on water quality from STORET and data on aquatic resources, to help identify water bodies which may not be meeting water quality standards.	OW. Jerry Laveck (202) 260-7771
CERCLIS 3	Database	Comprehensive Environmental Response, Compensation, and Liability Information System (CERCLIS) contains information on all aspects of hazardous waste sites from initial discovery to listing on the National Priorities List.	OERR. OSWER Superfund Automated Phone System (202) 260-8321 Margret Brown brown.margret@epamail.epa.gov.
ECOTOX	Database	The Ecotoxicology Database Retrieval System (ECOTOX) system represents an integration of AQUIRE, PHYTOTOX, AND TERROTOX, which are three existing EPA databases that contain ecotoxicity information for aquatic life, terrestrial plants, and wildlife respectively.	ORD. Christine Russom (218) 529-5000
EIMS	Database/Website	The Environmental Information Management System (EIMS) contains metadata about environmental data sets in a form searchable by source, location, and keywords. EIMS contains no taxonomic metadata, but stores metadata about collections of data, some of them biological.	OW. Bob Shepanek (202) 260-3255
Exams-II	Model	The Exposure Analysis Modeling System (EXAMS-II) is an interactive modeling system that allows users to specify and store the properties of chemicals and ecosystems, modify the characteristics of them via simple English-like commands and conduct rapid evaluations and error analyses of the probable aquatic fate of synthetic organic chemicals.	ORD. Model Coordinator (706) 546-3549

System	Application Type	Abstract	Contact Information
FGETS	Model	Food and Gill Exchange of Toxic Substances (FGETS) is a FORTRAN simulation model that predicts temporal dynamics of a fish's whole body composition [(ug chemical)/(g live weight fish)] of nonionic, nonmetabolized, organic chemicals that are bioaccumulated from either: (a) water only - which is the pre-dominant route of exchange during acute exposures, or (b) water and food jointly - which is more characteristic of chronic exposures.	ORD. Model Coordinator (706) 546-3549
FISHTEMP	Database	The National Compendium of Freshwater Fish and Water Temperature (FISHTEMP) is a computer database, containing historical fish distribution data with accompanying water temperature data from about 1930 to present for over 300 species of freshwater fish from 250,000 locations in the U.S.	ORD. John Eaton (218) 720-5557
GCSOLAR	Model	The Green Cross Solar (GCSOLAR) program is a set of routines that computes direct photolysis rates and half-lives of pollutants in the aquatic environment. The half-lives are calculated as a function of season, latitude, time-of-day, depth in water bodies, and ozone layer thickness.	ORD. Model Coordinator (706) 546-3549
LAMS	Database/Model	Lake Analysis Management System (LAMS) is a system of databases and models, with GIS capability, developed for the Great Lakes and watersheds.	ORD. Russell Kreis (313) 692-7615
LAST	Database	The Labor and Sample Tracking System (LAST) provides locally controlled storage of environmental sample analysis data. This data is stored until all lab test results for a sample on a site are complete, verified by the lab, and released to the STORET system.	Region 7 Kansas City. Dale Bates (913) 551-5183
PIN	Database	The Pesticide Information Network (PIN) is a menu-driven, interactive database containing current and historic pesticide information.	OPPTS. Leslie Davies-Hilleard (703) 305-7499
QUAL2E	Model	This software product can be used to study the impact of waste loads on instream water quality.	ORD. Model Coordinator (706) 546-3180
R2STREAM	Spreadsheet	Fish and macroinvertebrate assessments are conducted on streams and rivers located in New Jersey, New York, and Puerto Rico. All samples are collected by EPA and used for surface water quality assessments. Measurements of fish and macroinvertebrate community composition and relative abundance are used to calculate multi-metric indices.	Region 2 New York. Jim Kurtenback (732) 321-6695

System	Application Type	Abstract	Contact Information
ReachScan	Model/Database	ReachScan, the Probabalistic Dilution Model (PDM), and the Endangered Species Database (ESDB) have been integrated into one package for distribution. ReachScan is an integrated surface water modeling and database system used to: 1) estimate surface water concentrations at, and populations served by, drinking water utilities downstream from industrial facilities; 2) estimate concentrations in single or multiple reaches by simple dilution or using simple fate algorithms; 3) estimate the days of exceedance of a concern concentration in single or multiple reaches, the PDM-link; 4) determine the presence of endangered species or critical habitats in the county of the releasing facility, the ESDB-link; and 5) serve as a database for the identification of facilities and utilities.	OPPTS. Sidney Abel (703) 305-7346
RITZ	Model	The Regulatory and Investigative Treatment Zone Model (RITZ) is a useful tool for predicting fate and transport potentials of hazardous organic constituents contained in contaminated soils. It is a one dimensional, unsaturated flow and transport model. The model provides an estimate of the amount of each organic constituent which will be volatilized, transformed, leached, and retained in a defined zone of the soil. One of the principal features of RITZ is its ability to account for effects of the presence of oil in the waste-soil matrix. The model describes the subject site as consisting of two zones: the plow zone where the contaminated material is applied; and the transport zone where the contaminant is transported and transformed.	ORD. David Burden (405) 436-8606
RQ DB	Database	The database supports the regulatory development effort of the Response Standards and Criteria Branch within the Emergency Response Division (ERD) of the Office of Emergency and Remedial Response. The Reportable Quantities (RQ) database contains the names, regulatory synonyms, and chemical abstracts service registry numbers (CASRNs). It also contains chemical-specific data on each of the CERCLA hazardous substances and on the chemicals that ERD is considering for addition to the list of hazardous substances pursuant to the authority in Section 102(a) of CERCLA.	OSWER. Gerain Perry (202) 564-6704

System	Application Type	Abstract	Contact Information
SDWIS	Database	The Safe Drinking Water Information System (SDWIS/FED) maintains inventory and compliance data (violations and follow-up actions) reported by PRIMACY agents under the Public Water Supervision System(PWSS). It is used by Headquarters to provide quarterly reports to other components of EPA and to satisfy external reporting requirements. It is also used by Headquarters, Regions, and states to perform oversight. It currently contains compliance information from 1980 to the present and relates follow-up actions to specific violations.	OW. Towana Dorsey (202) 260-2805
STORET	Database	The Water STOrage and RETrieval (STORET) system assists State and EPA officials in making pollution control decisions by providing a capability to store, retrieve, and analyze water quality information. Data within STORET are used in support of 305(b) reporting requirements and trend analysis. Both ground water and surface water information are contained in the system and represent the results of ambient monitoring events at over 800,000 sites throughout the United States. The system has a set of analytical tools that can generate data listings, criteria reports, statistics, maps, and graphs. STORET also contains a biological taxonomy.	OW. Lee Manning (202) 260-6082
SURF	Website	The Index of Watershed Indicators section of the Surf Your Watershed (SURF) website (http://www.epa.gov/surf) contains a data layer for Aquatic/Wetland Species at Risk, as well as numerous other data layers related to watersheds.	OW. Karen Klima (202) 260-7087
SWS	Database	The Streamwalk (SWS) program is a public outreach system designed to compile observational data from volunteers in the field about stream conditions for streams they adopt. The database is designed to be used by volunteers with or without PC's. Streamwalk observational data is transferred to a central system at the region for compilation of data.	Region 10 Seattle. Laurie Mann (206) 553-1583

System	Application Type	Abstract	Contact Information
TRIS	Database	The Toxics Release Information System (TRIS) contains all non-Trade Secret data submitted to EPA under the Act for chemicals and chemical categories listed by the Agency. Data include chemical identity, amount of on-site users, releases and off-site transfers (including POTW), on-site treatment, and minimization/prevention actions. The EPA internal system (available to authorized users of the ES 9000) is as described above. A public-access system is provided by the National Library of Medicine through TOXNET. TRIS contains only two data elements that include biological treatment and biological degradation which may not carry any taxonomic information.	OPPTS. Ruby Boyd (202) 260-8387
Multi-Jurisdictional Database	Database	The Multi-Jurisdictional database is being developed by The Nature Conservancy, for use by EPA and other funding agencies in identifying species diversity within ecosystems. The data will be drawn from State Heritage Program data, which includes plant and animal species occurrences in the 50 States and some U.S. territories. The data file will include threatened and endangered species data that will be searchable via the Internet at the County and watershed level. More specific locational information will be available, but only via secure access, to ensure security of information on sensitive habitats. Widely accepted plant and animal taxonomies are being used in the identification of species. Although EPA is not the owner of this database, they have, in part, sponsored its development.	OIRM. Jacques Kapuscinski (202) 260-3129
EMAP	Geographic Information System	The Environmental Monitoring and Assessment Program (EMAP) has evaluated its progress since 1989 and the recommendations of 20 peer reviews. The program retains its goal to: Monitor the condition of the Nation's ecological resources to evaluate the cumulative success of current policies and programs and to identify emerging problems before they become widespread or irreversible. The strategy for EMAP is based on three principles. First, pursue all tiers in the monitoring framework (i.e., Index Sites, Geographic Surveys, and Landscape Monitoring). Second, focus the next three years on the research and demonstration necessary to provide the scientific credibility for the monitoring network. Third, based on the knowledge of the science necessary for success, build the national network from the bottom up, starting with effective existing networks and add to them where gaps exist.	RTP. Laura Jackson (919) 541-3088

APPENDIX B

Data Elements from EPA and Other Information Systems
Using Biological Information

Data Elements for EPA and Other Information Systems Using Biological Information

ITIS Elements	FGDC/NBII Elements	STORET	AQUIRE	ECOTOX	The Nature Conservancy	SDWIS	TRIS
Kingdom Name	Kingdom						
Vernacular_Name	Applicable Common Names					Name	
Taxonomic_Unit_Types	General Taxonomic Information	Taxon_Rank_Code					
Taxonomic_Credibility_ Rating	Taxonomic Completeness				Taxonomy Comments		
Rank_name	Kingdom, Division-Phylum, Class, Order Family, Genus, Species		Phylum, Class, Family, Subclass, Order		Class, Subspecies, Families, Superfamily, Phylum, Subphylum, Order		
Geographic_value	Geographic Extent						
Source_type	Classification System or Authority						
Source	Classification System Citation						
	Methodology	Field_Activity_Type_Name					Treatment_Method _Code
	Methodology Type	Community_Type					
	Methodology Keyword	Type_Name					Efficiency_Est
	Methodology Identifier	TBD_Area_Description_Text					
Taxonomic Serial Number		Species_Number					

ITIS Elements	FGDC/NBII Elements	STORET	AQUIRE	ECOTOX	The Nature Conservancy	SDWIS	TRIS
				Aquatic Toxicity Effects Data			
				Terrestrial Plant Effects Data			
				Terrestrial Animal Effects Data (TERRETOX)			
					Vertebrates Higher taxonomy source		
					Invertebrates Higher taxonomy source		

APPENDIX C

FGDC/NBII Data Elements

FGDC/NBII Data Elements

FGDC/NBII Elements	FGDC/NBII Definitions
Taxonomy	Information on the taxa (one or more) included in the data set, including keywords and taxonomic coverage information.
Taxonomic Keywords	Common-use words or phrases describing the taxonomy covered by the data set.
Taxonomic Coverage	Information about the range of taxa addressed in the data set or collection. It is recommended that one provide information to a level which reflects the data set or collection being documented.
General Taxonomic Information	A description of the range of taxa addressed in the data set or collection. For example, all vascular plants were identified to family or species, mosses, and lichens were identified as moss or lichen.
Specific Taxonomic Information	Specification of the taxa addressed in the data set or collection.
Kingdom	Specification of the Kingdom name.
Division-Phylum	Specification of the Division/Phylum name.
Class	Specification of the Class name.
Order	Specification of the Order name.
Family	Specification of the Family name.
Genus	Specification of the Genus name.
Species	Specification of the Species name, including subspecies, variety name, and author citation (with date as appropriate).
Applicable Common Names	Specification of applicable common names. These common names may be general descriptions of a group of organisms, if appropriate (e.g., insects, vertebrate, grasses, waterfowl, vascular plants, etc.)
Geographic Extent	The geographic areal domain of the data set. This field is applicable if the data set has any relationship to a geographic location.
Taxonomic System	Documentation of taxonomic sources, procedures, and treatments.
Classification System or Authority	Information about the classification system or authority used.
Classification System Citation	A citation for the classification system or authority used, this might include monographs (e.g., a regional flora) or on-line data sets (e.g., the USDA PLANTS database), etc.
Classification System Modifications	A description of any modifications or exceptions make to the classification system or authority used.
Identification Reference	Information on any non-authoritative materials (e.g., field guides) useful for reconstructing the actual identification process.
Identifier	Information about the individual(s) responsible for the identification(s) of the specimens or sightings, etc.
Taxonomic Procedures	Description of the methods used for the taxonomic identification. Could include specimen processing, comparison with museum materials, keys and key characters, chemical or genetic analyses, etc.

FGDC/NBII Elements	FGDC/NBII Definitions
Taxonomic Completeness	Information concerning the proportions and treatment of unidentified materials (i.e., materials sent to experts and not yet determined); estimates of the numbers, importance, and identities of misidentifications, uncertain determinations, synonyms or other incorrect usages; taxa not well treated or requiring further work; and expertise of field workers.
Vouchers	Information on the types of specimen, the repository, and the individuals who identified the vouchers.
Specimen	A word or phrase describing the type of specimen collected.
Repository	Information about the curator or contact person and/or agency responsible for the specimens.
Methodology	Information about a single step of field and/or laboratory work.
Methodology Type	Is field methodology or laboratory methodology being documented.
Methodology Identifier	Keywords or phrases summarizing the field or laboratory methods used.
Methodology Keyword Thesaurus	Reference to a formally registered thesaurus or a similar authoritative source of methodology keywords.
Methodology Keyword	The name of a method used in the field or laboratory work.
Methodology Description	Equivalent to "Material and Methods" in a journal article. Describe the physical methods used to gather data, the experimental design, sample frequency, treatments or strata, statistical and spatial design of the sampling, and sample completeness, representatives, and biases. For example, in a bird survey, relevant elements would include the methods used to detect species occurrences (casual sightings, transects, focal point surveys, vocalizations, or mist nets).
Methodology Citation	Information referencing the methods used.

APPENDIX D

ITIS Data Elements

ITIS Data Elements

ITIS Elements	ITIS Definitions
Usage	Current standing of an occurrence of Taxonomic Units.
Unacceptability_reason	The cause for an occurrence of Taxonomic Units being identified as not accepted/invalid under the usage element.
Taxonomic_Unit_Types	Defines the levels associated with the taxonomic hierarchical structure and establishes the rank order for an occurrence of the Taxonomic Units.
Taxonomic Serial Number	Taxonomic Serial Number. The unique identifier for an occurrence of Taxonomic Units.
Parent_TSN	The taxonomic serial number for the direct parent of the subject occurrence of Taxonomic Units.
Taxonomic_Credibility_Rating	A subjective rating designation as determined by the Taxonomic Work Group reflecting the level of review and the perceived level of accuracy for an occurrence of Taxonomic Units and its associated attributes.
Taxon_Authors_Lookup	Reference to authors of taxons for the Plant kingdom, and authors and dates of taxons for the Animal kingdom.
Taxon_Author	For authorities associated with the animal kingdom, the author(s) and the actual date of the publication in which the author(s) first described the taxon are required. For authorities associated with the plant kingdom, only author(s) is/are required. Note: Initially, authors/dates may not be available for all records.
Taxon_Author_ID	A unique identifier for the author(s) of a taxonomic name.
Kingdoms	The highest rank in the taxonomic hierarchical structure.
Kingdom_name	The label associated with the highest level of the taxonomic hierarchy structure.
Kingdom_id	A unique identifier for the highest level of the taxonomic hierarchy structure.
Rank_name	The label associated with the specific level of a taxonomic hierarchy.
Rank_ID	A unique identifier for a specific level within the taxonomic hierarchy.
Direct_Parent_Rank_ID	The unique identifier for the rank of the closest parent of an occurrence of Taxonomic Units as defined by the kingdom's rank rules.
Required_parent_rank_id	The unique identifier for the closest, required parent of an occurrence of Taxonomic Units as established by the respective kingdom's rank rules.
Origin	Indication of whether an occurrence of Taxonomic Units is native and/or introduced to a U.S. jurisdictional unit.
Jurisdiction_Value	Label signifying a U.S. jurisdictional unit as defined by the Taxonomy Work Group and Canada.
Geographic_value	Label given to a geographic division as identified by the Taxonomic Work Group.
Vernaculars	Common names associated with an accepted occurrence in Taxonomic Units.
Vernacular_Name	A common name associated with an occurrence of Taxonomic Units.
Language	Native language from which the vernacular name originates; e.g., American, English, Russian, Spanish, etc.

ITIS Elements	ITIS Definitions
Review_reason	An indication of the purpose for which a review of an occurrence of Taxonomic Units was undertaken.
Reviewer	The person to whom the Taxonomic Work Group delegated the task of reviewing an occurrence of Taxonomic Units.
Comments	A mechanism for recording history or detail deemed important for an occurrence(s) of Taxonomic Units.
Comment_id	The unique identifier for remarks associated with an occurrence of Taxonomic Units.
Change_Tracks	An audit mechanism for tracking changes made to the taxonomy, nomenclature and/or attributes of an occurrence of Taxonomic Units.
Change_track_id	The unique identifier assigned to a change made to an occurrence of Taxonomic Units.
Old_TSN	The taxonomic serial number associated with an occurrence of Taxonomic Units, the changes to which result in the addition or other change to the subject occurrence of Taxonomic Units.
Source_type	The designation of the kind of supplier providing information to the ITIS (other than a person or publication); e.g., database.
Source	The name of the supplier of information, other than a person or publication, to the ITIS. Examples include NODC, PLANTS, BIOTA, ZOO Record, etc.
Source_comment	Remarks associated with the provider of information to the ITIS (other than a person or publication).
Unit_ind1	Indicator of an occurrence of a plant hybrid at the generic level.
Unit_name1	The singular or first scientifically accepted label for an occurrence of Taxonomic Unit.
Unit_ind2	A category indicator positioned between the first and second parts of a binomial/polynomial taxonomic name.
Unit_name2	The second part of a scientifically accepted label for binomial/polynomial occurrence of a Taxonomic Unit.
Unit_ind3	A category indicator located within a polynomial taxonomic name.
Unit_name3	The third portion of a scientifically accepted label for a polynomial occurrence of Taxonomic Units.
Unit_ind4	A category indicator located within a polynomial taxonomic name.
Unit_name4	The fourth part of a scientifically accepted label for a polynomial occurrence of Taxonomic Units.
Unnamed_taxon_ind	Indicator for occurrences of Taxonomic Units whose names represent unnamed taxa and are therefore not standard taxonomic units.

APPENDIX E

Proposed Set of Data Elements
for
Biological Taxonomy Identification

Proposed Set of Data Elements for Biological Taxonomy Identification

Data Element Name	EDR Data Element Number	Data Element Definition
ITIS Taxonomic Serial Number	20727	The unique serial number assigned by ITIS to a biological entity.
EPA Biological Identification Number	20728	The unique, non-intelligent number assigned to a biological entity by the Environmental Protection Agency.
Biological Systematic Name	20729	The name assigned to a biological entity by a classification identification system.
Biological Kingdom Name	20730	The systematic name that represents the Kingdom.
Biological Division/Phylum Name	20731	The systematic name that represents the Division/Phylum.
Biological Subdivision/Subphylum Name	20732	The systematic name that represents the Subdivision/Subphylum.
Biological Class Name	20733	The systematic name that represents the Class.
Biological Subclass Name	20734	The systematic name that represents the Subclass.
Biological Infraclass Name	20735	The systematic name that represents the Infraclass.
Biological Order Name	20736	The systematic name that represents the Order.
Biological Suborder Name	20737	The systematic name that represents the Suborder.
Biological Family Name	20738	The systematic name that represents the Family.
Biological Subfamily Name	20739	The systmatic name that represents the Subfamily.
Biological Genus Name	20740	The systematic name that represents the Genus.
Biological Species Name	20741	The systematic name that represents the Species.
Biological Subspecies Name	20742	The systmatic name that represents the Subspecies.

Data Element Name	EDR Data Element Number	Data Element Definition
Biological Systematic Name Context	20743	The name of the classification system used to assign of the systematic name to a biological entity.
Biological Vernacular Name	20744	The vernacular name associated with an occurrence of biological entity.
Biological Vernacular Name Context	20745	The name of the source of a vernacular name of a biological entity.
Biological Synonymous Name	20746	The name that represents one of two or more names applied to a single biological entity.
Biological Synonymous Name Context	20747	The name of the source of a synonymous name for a biological entity.
Program System Biological Name	20748	The name applied to a biological entity in a program system.
Program System Biological Name Context	20749	The name of a program system that is the source of a biological entity name.
Biological Group Name	TBD	The name of a collection of biological entities that are related.
Biological Group Text	20750	The text describing the relationship of the biological entities in a group.
Biological Citation Text	20751	The text that describes the official references where requirements for monitoring a biological entity are defined.